

# Protocolo TGP

## Introducción:

TGP es un protocolo basado en texto para el armado del árbol de trabajo del sistema de análisis de logs. Se utilizó un protocolo basado en texto porque brindaba muchas más ventajas que desventajas en cuanto a implementación y performance. Principalmente, la decisión se basó en que un protocolo basado en texto es human-readable y por ende más fácil de testear y entender por un humano, lo cual facilita enormemente la implementación y testeo del protocolo. Además, al ser un protocolo que se usa para inicializar el árbol de trabajo, sabemos que [el protocolo] se usará sólo en la etapa inicial del funcionamiento del sistema, con algún agregado ocasional si se agrega un worker on-the-fly. Por ende, decidimos que la ventaja en performance que se ganaba usando un protocolo orientado a bytes (al utilizarse menos bytes para transmitir la misma información) no tenía peso suficiente como para importar más que la comodidad de uso y desarrollo de un protocolo basado en texto.

## Objetivo:

Este protocolo se propone para el armado del árbol de trabajo de workers/directors. El mismo incluye el descubrimiento de directores por parte de los workers, el ofrecimiento de los directores para que el worker se asigne a su grupo de trabajo, y el acuerdo que hacen el worker y el director para ligarse en un grupo de trabajo. Dada la similitud que tiene el mecanismo de armado del árbol de trabajo con el de asignación de IPs en una red, se tomó como base y modelo el protocolo DHCP.

Ejemplo de un intercambio de mensajes típico entre un worker y un director:

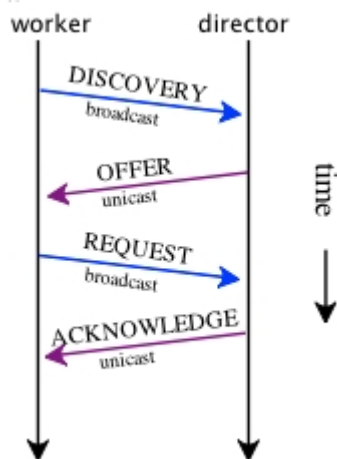


Imagen basada en una imagen similar explicativa del protocolo DHCP en [wikipedia](https://es.wikipedia.org/wiki/Protocolo_DHCP).

## Mensajes:

Estructura General:

```
COMMAND
parameters
```

donde COMMAND es uno de los siguientes: TGPDISCOVER, TGPOFFER, TGPREQUEST, TGPACK y

parameters son pares clave-valor siguiendo la sintaxis "key: value\n".

Cada mensaje tiene sus parámetros específicos. Si alguna de las partes recibe un mensaje con parámetros erróneos u otra estructura de mensaje, puede ignorarlo.

**TGPDISCOVER:** un worker que no tiene director asociado puede enviar un paquete TGPDISCOVER para encontrar directores a los cuales ligarse.

El mensaje TGPDISCOVER contiene un único parámetro opcional `group` que es el grupo del que se quiere descubrir el director. Si este parametro no está seteado, el TGPDISCOVER se interpreta como que el worker está buscando descubrir al director de cualquier grupo y que está dispuesto a trabajar para cualquier director. En caso de estar seteado, el parámetro `group` se interpreta cómo que el único director al que este worker se quiere ligar es al director del grupo con número `group`. Sólo el director de este grupo podrá responder a este mensaje. En caso de no recibir respuesta, el worker podrá intentar nuevamente enviando un nuevo mensaje TGPDISCOVER (y si, por ejemplo, el director del grupo pedido no respondió, puede omitir este campo para unirse a un grupo cualquiera)

Ejemplos de mensajes TGPDISCOVER válidos son:

```
TGPDISCOVER
group: 2
```

o

```
TGPDISCOVER
```

**TGPOFFER:** un director que recibe un TGPDISCOVER le responde al worker diciéndole que su grupo lo acepta como trabajador.

Cuando un director recibe un DHCPDISCOVER de un worker, se fija si él es el director del grupo especificado (si es que se especificó un grupo). En caso negativo, ignora el mensaje, y en caso positivo, si es capaz de aceptar al nuevo worker dentro del grupo de trabajo, le reserva un lugar y envía un mensaje TGPOFFER diciéndole de qué grupo se trata en el parámetro obligatorio `group`. Si el director no es capaz de aceptar este nuevo trabajador, simplemente ignora el TGPDISCOVER. En caso de no recibir respuesta (TGPREQUEST) del worker, el director puede liberar el lugar reservado para el worker.

Ejemplos de mensajes TGPOFFER válidos son:

```
TGPOFFER
group: 2
```

o

```
TGPOFFER
group: 7
```

**TGPREQUEST:** un worker, al recibir TGPOFFER de uno (o varios) directors, decide responderle a uno de ellos pidiéndole incorporarse a su grupo de trabajo.

Cuando un worker recibe un TGPOFFER de uno o varios directors, debe responder a lo sumo a uno, indicándole que quiere incorporarse a su grupo de trabajo. Este mensaje contendrá un parámetro obligatorio `group` que indicará a qué director está dispuesto a ligarse. El mensaje es broadcast, y por ende si los otros directores que le habían ofrecido pertenecer a su grupo reciben este mensaje, pueden liberar el espacio que le habían reservado, ya que saben que está intentando unirse a otro grupo. El mensaje TGPREQUEST también contiene como parámetros obligatorios `host` y `port`, que son el ip y el puerto en el que el worker atiende pedidos de procesamiento de logs (detallados en el protocolo WDP). Nótese que la ip y el puerto en el que atiende estos pedidos no tienen nada que ver con la ip y el puerto en el que atiende los mensajes TGP.

Ejemplos de mensajes TGPREQUEST válidos son:

```
TGPREQUEST
group: 2
host: 195.165.12.9
port: 5132
```

o

```
TGPREQUEST
group: 1
host: 10.0.0.66
port: 6655
```

**TGPACK:** un director acepta el pedido TGPREQUEST de un worker para unirse a su grupo de trabajo e inicia una conexión TCP para la distribución del trabajo.

Este mensaje indica que el director se da por enterado de que el worker ha entrado en su grupo de trabajo, y por lo tanto confirma la toma del lugar del nuevo worker en el grupo. Luego de enviar este mensaje, el director se conecta al worker mediante el endpoint provisto por el mensaje TGPREQUEST anterior. El worker, al recibir este mensaje, sabe que el director lo tomó para su grupo de trabajo, y deja de buscar un director nuevo. Si al intentar conectarse por TCP, el director se da cuenta que el worker aún no está escuchando en dicho puerto, puede retransmitir el TGPACK.

El único ejemplo de mensaje TGPACK válido es:

```
TGPACK
```

